Nucleic Acids Research, 1993, Vol. 21, No. 6 1409–1417

Duplicated KOX zinc finger gene clusters flank the centromere of human chromosome 10: evidence for a pericentric inversion during primate evolution

Alan Tunnacliffe*, Lizhi Liu+, Julie K.Moore, Margaret A.Leversha[1], Michael S.Jackson, Laura Papi§, Malcolm A.Ferguson-Smith[1], Hans-Jürgen Thiesen[2] and Bruce A.J.Ponder
CRC Human Cancer Genetics Research Group and [1]Human Molecular Genetics, Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QP, UK and [2]Basel Institute for Immunology, Grenzacherstrasse 487, CH-4058 Basel, Switzerland

Received December 24, 1992; Revised and Accepted February 11, 1993 EMBL accession nos¶

ABSTRACT

Two related zinc finger (ZNF) gene clusters from the pericentromeric region of human chromosome 10, defined by cDNAs of the KOX series, have been cloned in yeast artificial chromosomes (YACs). The two clusters evolved by duplication of an ancestral gene cluster before the divergence of the human and great ape lineages. Included in cluster A are the ZNF gene sequences *ZNF11A*, *ZNF33A*, and *ZNF37A*, while cluster B comprises the related sequences *ZNF11B*, *ZNF33B* and *ZNF37B*. Genes from both clusters are expressed: cDNAs KOX2, KOX31 and KOX21 derive from *ZNF11B*, *ZNF33A* and *ZNF37A*, respectively. Further YACs have been isolated which link *ZNF11A* and *ZNF33A* to another gene, *ZNF25*, defined by cDNA clone KOX19. Therefore *ZNF25* also forms part of cluster A, but has no counterpart in cluster B. Surprisingly, the KOX ZNF gene clusters are located on opposite sides of the centromere: cluster A maps to 10p11.2, while cluster B is in 10q11.2. This suggests the occurrence during primate evolution of a previously undescribed pericentric inversion subsequent to the cluster duplication. The evolution of this subset of KOX ZNF genes has therefore involved three types of genetic event: local gene duplication, gene cluster duplication, and chromosome rearrangement.

INTRODUCTION

The pericentromeric region of human chromosome 10, 10p11.2–q11.2, has aroused considerable interest since it was shown to contain the genes for the inherited cancer syndromes multiple endocrine neoplasia (MEN) types 2A and 2B (1,2). To facilitate the cloning of the MEN 2 genes, physical mapping and cloning projects in the pericentromeric region of chromosome 10 were initiated. Relatively few cloned genes have been assigned to 10p11.2–q11.2, but these include several zinc finger (ZNF) protein genes (3–5). ZNF proteins contain 'finger' motifs consisting of approximately 30 amino acids which coordinate a zinc ion, and were first described in the Xenopus transcription factor IIIA (6,7), but have been shown to be present in species as diverse as Drosophila and man. The ZNF gene family may consist of several hundred members in the human genome (8,9), and includes examples which have been implicated in the control of cell growth and differentiation (10,11).

We have begun to characterise several of the ZNF genes located in the region 10p11.2–q11.2 which are defined by the KOX series of cDNAs (9). The KOX2 cDNA clone has been shown to hybridise to two loci in the pericentromeric region of human chromosome 10, one of which (*ZNF11A*) was assigned to 10p11.2–q11.2, while the second (*ZNF11B*) could be mapped more precisely, to 10q11.2 (5). Because both loci hybridise with the KOX2 probe under stringent conditions, it was hypothesised that *ZNF11A* and *ZNF11B* are highly related and have arisen by an evolutionarily recent duplication. This hypothesis was tested and confirmed here by cloning both *ZNF11* loci and comparing their sequences. Furthermore, we have also explored at the molecular level the question of clustering of ZNF genes, since many such genes have been shown to share similar chromosomal locations using somatic cell hybrid and cytogenetic mapping (eg. refs. 12–14). Finally, we have shown that our results have implications for the evolution of human chromosome 10 and for the organisation of genes around its centromere.

MATERIALS AND METHODS

Yeast artificial chromosomes (YACs)

The Washington University YAC library (15) was screened by hybridisation of high density filters with the KOX2 (9) cDNA probe. Primary pools of the CEPH MkI YAC library were screened by PCR using primers K19A and K19B for *ZNF25*,

* To whom correspondence should be addressed

Present addresses: +Department of Neurobiology, Institute of Animal Physiology, Genetics Research, Babraham, Cambridge CB2 4AT, UK and §Department of Clinical Physiopathology, Section of Medical Genetics, Viale Pieraccini 6, 50134 Firenze, Italy

¶X68684–X68689 (incl.), X69115 and X69116

and primers 4113 and 4114 for *ZNF33* sequences (see below). Secondary screens were performed by CEPH. Protocols for yeast handling, including agarose plug preparations and PFGE, have been described previously (16,17). PFGE was carried out using a BioRad (Richmond, CA) CHEF DRII system as described (18). YAC yA136E7 (~400 kb) probably contains a co-ligated insert, since FISH experiments (see below) initially gave signals on 18q11.2 in addition to those on 10q11.2. In subsequent experiments, a signal on 1q25 was also observed, which probably indicates the presence of a second YAC in the yeast clone. Clone yA4C5 (220 kb) maps to 10p11.2, but the yeast isolate also contains a second contaminating YAC (~330 kb) which gives FISH signals on 3q21.

Polymerase chain reaction (PCR)

All novel oligonucleotide primers were designed to amplify target DNAs under standard cycling conditions of 94°C 1 min, 55°C 1 min, 72°C in the buffer of ref. 19. For KOX2-related sequences, these primers were AK2A (5′ AAATGCTGACTAAGGAACAAGGT 3′; positions 110–132 in Figure 2a) and AK2B (5′ CCCTGTGTGTGTTCTCTGATG 3′; positions 1086–1066 in Figure 2a) giving a fragment of 977 bp; for KOX31-related sequences, 4113 (5′ GCCATAAGTCAGCTCTAATTG 3′; positions 55–75 in Figure 2b) and 4114 (5′ GCTTTACATTCCACAATATACATC 3′; positions 904–881 in Figure 2b) giving a fragment of 850 bp. For KOX21-related sequences, the oligonucleotide primers used were K21FOR (5′ TGAAAAGTCAACCCTTACTCAAC 3′; positions 56–78 of the KOX21 cDNA sequence: accession number X69115) and K21REV (5′ GGCAAAATGTTTCCTATATTCAT 3′; positions 974–952 of X69115) giving a fragment of 919 bp.

Sequencing

The United States Biochemical (Cleveland, OH) Sequenase kit was used throughout. The cDNA clones were sequenced from both ends using standard forward and reverse primers. Subsequently, specific primers as follows were used to complete the full length sequence: for KOX2, primers 748 (5′ CAAGGTAATGTAATAGGAATA 3′; positions 127–147 of Figure 2a), ED1/K2 (5′ GCATGATGAAACTCATAC 3′; positions 318–335), ED4/K2 (5′ TATAATGAATTTGGGAGA 3′; positions 520–537), and ED7/K2 (5′ AAGTTCTTATGTGTGAAG 3′; positions 619–636); for KOX31, primers 4113 NEST (5′ ATTGTACATCAGAGAACCC 3′; positions 72–90 of Figure 2b), ED2/K31 (5′ GTGGAAAAATCTTTTACC 3′; positions 292–309), ED5/K31 (5′ TGGATATTAGAAATTTCC 3′; positions 454–471); for KOX21, primers K21NEST (5′ TTACTCAACATCAAAGAACG 3′; positions 70–89 of X69115), ED3/K21 (5′ GGAAGTCATTCTCTGAGA 3′; positions 295–312), and ED6/K21 (5′ GAAAATCATTCTATGTTA 3′; positions 631–648). For direct sequencing of PCR fragments derived from YACs and cDNA clones, a biotinylated 3′ PCR primer was used (AK2B for KOX2; 4114 for KOX31; K21REV for KOX21). The biotinylated fragments were purified with Dynatech magnetic beads coated with avidin, made single stranded, and sequenced using the above-listed primers, while attached to the beads as described (20). Gel readings were assembled into contigs using the IBI AssemblyLign program on Apple Macintosh computers. Sequence analysis was with the Macvector (IBI) and GCG packages. A partial sequence for the KOX2 cDNA was published previously (accession no. X52333) which contains two errors, corrected in this paper. The sequences presented in this paper have the accession numbers X68684–X68689, and X69115–6 (EMBL database). Divergence between related pairs of *ZNF11* and *ZNF33* sequences was calculated according to the two-parameter method of ref. 21.

Somatic cell hybrids

Hybrids TG3 and TK2 have been described (22). The subclones R342A4-B14, R342A4-B15 and R342A4-C3 were obtained from the human chromosome 10-only hybrid R342A4 (23) following spontaneous deletion or BUdR-induced mutagenesis (24). They were analysed by chromosome painting (Figure 4), and by PCR-based marker analysis which showed them to lack all 10q markers tested. Hybrid JF1-23B13 (25) was obtained from a fusion of hamster cell line Ade6 (CHO proline auxotroph) and human lymphoblasts carrying a balanced t(10;21)(p11.2;q22.3) translocation (cell line GM06135; NIGMS Human Genetic Mutant Cell Repository, Camden, NJ). Cytogenetic analysis showed the hybrid to carry the translocation chromosome containing region 10pter-p11.2 and no other chromosome 10 material, although other human chromosomes were also present.

Filter hybridisations

Standard protocols (26) were followed for restriction digests, agarose gel electrophoresis and Southern blotting to nylon membranes (Zetaprobe, BioRad; or Hybond N, Amersham Life Sciences, Amersham, UK). Hybridisations resulting in the autoradiographs of Figures 1 and 5 were carried out in 3×saline sodium phosphate (SSPE), 10% dextran sulphate, 0.5% sodium dodecyl sulphate (SDS), Denhardt's and 100 μg/ml sonicated salmon sperm DNA at 65°C overnight. The final filter washes were in 0.2×saline sodium citrate (SSC), 0.1% SDS at 65°C. Probes were labelled according to ref. 27: inserts from KOX cDNAs (9) were excised and purified by two rounds of agarose gel electrophoresis before labelling. For the zoo blot of Figure 3, some DNAs were obtained from cell lines: human (4502), chimpanzee (GM03452; NIGMS, Camden), orangutan (EB185(JC)); ECACC, Porton Down), African green monkey (CV-1), dog (MDCK) and mouse (Rag). Other samples were purchased from Clontech (Palo Alto, CA): cow, pig, rabbit and chicken. The remainder were gifts from Drs W.Amos and J.Pemberton (Genetics, Cambridge).

Chromosome painting/FISH

Human metaphase chromosome spreads were obtained from PHA-stimulated normal lymphocytes using standard procedures (28). 5-Bromodeoxyuridine (10μg/ml) was incorporated during the final 5 hours of culture. Biotinylated YAC and hybrid DNA was obtained by Alu-PCR using BK33 (29) or 517 (19) primers. The PCR products were ethanol-precipitated and 1 μg of DNA was digested with DNase I to produce fragments of 200–500 bp. The recovered pellet was dissolved in 10 μl TE buffer. Biotinylated probe DNAs were localised on R-banded normal male metaphase chromosomes by FISH according to procedures described elsewhere (Leversha et al., in preparation). The probe mix consisted of 0.5–1 μl labelled DNA, 1.5 μl sonicated placental DNA (10 mg/ml, Sigma) and 13 μl hybridisation buffer. Hybridisation was performed at 42°C overnight and the hybridised probe was detected using avidin-FITC, amplified by biotinylated anti-avidin and a further layer of avidin-FITC (30). The slides were counterstained with a mixture of 0.8 μg/ml DAPI and 0.4 μg/ml propidium iodide in buffered glycerol (31) and

viewed with a Nikon Optiphot epifluorescence microscope. Images of banded chromosomes with superimposed FITC signal were collected using a MRC Lasersharp 600 confocal imaging system (BioRad).

RESULTS

Two unlinked clusters of related ZNF genes in chromosome region 10p11.2−q11.2

Two *ZNF11* loci, which both hybridise to a cDNA probe KOX2 (9), are located on opposite sides of a translocation breakpoint in 10q11.2 (5). They have been separated in a pair of somatic cell hybrids (TG3 and TK2; 22), each of which carries one of the reciprocal translocation chromosomes. Figure 1a shows the hybridisation pattern after Southern blotting of DNA from hybrids TG3 (containing 10pter−q11.2) and TK2 (containing 10q11.2−qter), compared to controls. *ZNF11A*, mapping above the translocation breakpoint, is represented by a 4.0 kb EcoRI fragment, while *ZNF11B* is seen as a band at 1.75 kb which maps below the breakpoint. In order to investigate the relationship of these genes to each other, yeast artificial chromosomes (YACs) were isolated by screening with the KOX2 cDNA probe. Two YACs, yA136E7 and yA4C5, were analysed in most detail. Digests of YAC DNA, probed with KOX2, show that yA4C5 contains the 4.0 kb EcoRI fragment from the *ZNF11A* locus, while yA136E7 contains the 1.75 kb EcoRI fragment corresponding to *ZNF11B* (Figure 1b). KOX2 also hybridises to a large but distinguishable BamHI fragment in each YAC (Figure 1c). Since yA136E7 and yA4C5 only contain a single KOX2-hybridising fragment in each case, it is unlikely that the two *ZNF11* loci are closely linked in the genome.

Four further ZNF cDNAs (KOX15, KOX19, KOX21 and KOX31; 9) have been cloned whose genes map to 10p11.2−q11.2 (3,5,12), and the possibility arises of a cluster of related genes in this area of the genome. The respective cDNAs were therefore hybridised against yA136E7 and yA4C5 DNAs. No hybridisation was detected with KOX15 (gene symbol *ZNF22*) or KOX19 (*ZNF25*) probes. However, both YACs contain sequences homologous to KOX21 (*ZNF37*) and KOX31 (*ZNF33*) (Figure 1d, e), showing that there are linked *ZNF37* and *ZNF33* sequences at each *ZNF11* locus, and suggesting clustering of ZNF genes. As with the KOX2 probe, the hybridisation patterns obtained with KOX21 and KOX31 are clearly different for each YAC, implying that although each *ZNF37* and *ZNF33* sequence is related, they in fact represent distinct genes. To distinguish these sequences, we have provisionally termed those linked to *ZNF11A* in 10p11.2−q11.2 *ZNF37A* and *ZNF33A*, while those linked to *ZNF11B* in band 10q11.2 are named *ZNF37B* and *ZNF33B*. These KOX ZNF gene clusters are, in turn, called cluster A and cluster B, respectively.

The KOX31 cDNA hybridises to genomic BamHI fragments of identical size to those detected with the KOX2 probe (Figure 1c, e), suggesting close linkage of the *ZNF11* and *ZNF33* sequences in each YAC. The *ZNF11* and *ZNF33* sequences are

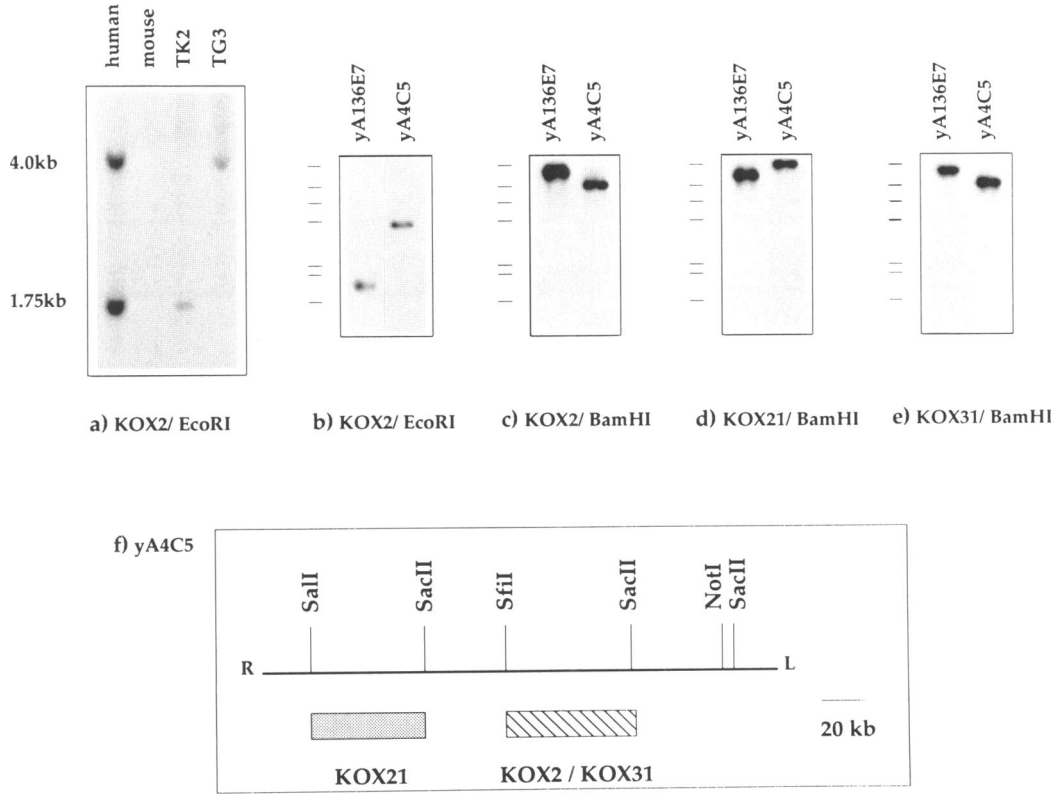

Figure 1. Southern hybridisation analysis of ZNF genes. a)−e) are filter hybridisations where genomic (a) or YAC (b−e) DNAs are probed with the indicated cDNAs. In **(a)**, hybrid TK2 contains 10q11.2−qter, and hybrid TG3 contains 10pter−q11.2. In **(b)−(e)**, marker positions (23, 9.4, 6.6, 4.4, 2.3, 2.0 and 1.3 kb) are represented by horizontal lines. **(f)** represents a PFGE-constructed long-range restriction map of YAC yA4C5 with regions hybridising to cDNA probes indicated. R and L denote right and left ends of the pYAC4 vector.

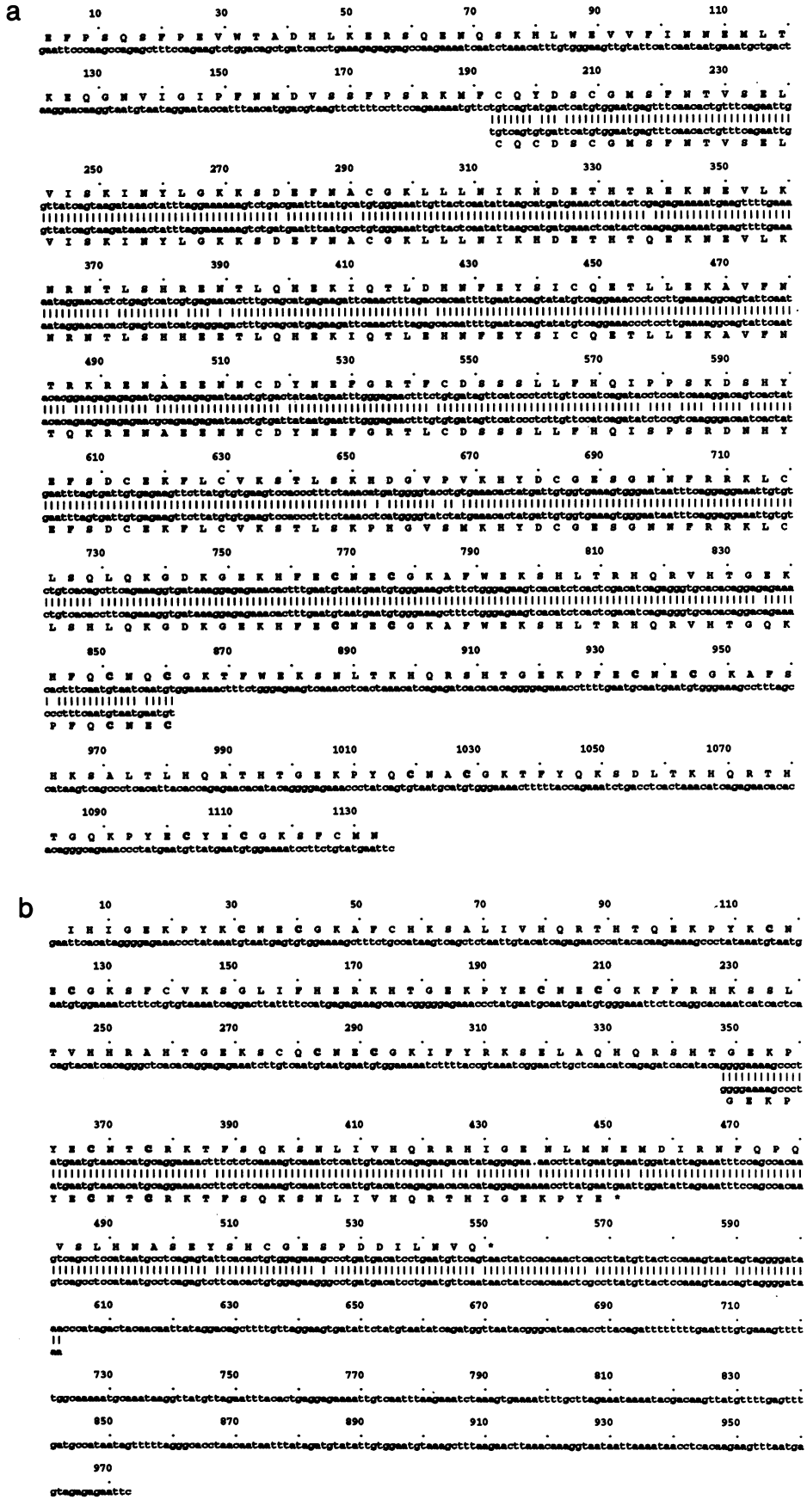

Figure 2. **(a)** Sequence of the KOX2 cDNA (top line) and part of the related sequence at *ZNF11A* in 10p11.2 (bottom line). **(b)** Sequence of the KOX31 cDNA (top line) and part of the related sequence at *ZNF33B* in 10q11.2 (bottom line). Nucleotide sequence is shown in lower case, with corresponding predicted amino acid sequence denoted in upper case (single letter code). The first two cysteines of each ZNF motif are shown in outline. Nucleotide matches are given as vertical bars.

clearly distinct, however, since other enzymes give different fragment sizes (12; and unpublished results). The long-range restriction map of one of the ZNF gene YACs, yA4C5, was constructed by pulsed field gel electrophoresis (PFGE), and is consistent with a close association of *ZNF11A* and *ZNF33A* sequences, since they are not separated by the restriction enzymes used (Figure 1f). In contrast, *ZNF37A* sequences are separated from *ZNF11A* and *ZNF33A* sequences by at least 35 kb in yA4C5.

In summary, there are clearly two distinct loci in the pericentromeric region of chromosome 10, both containing sequences which hybridise to three ZNF gene cDNAs. These sequences are organised as clusters: cluster A consisting of *ZNF11A*, *ZNF33A* and *ZNF37A*, while cluster B comprises *ZNF11B*, *ZNF33B* and *ZNF37B*. This implies that an ancestral duplication event occurred during the recent evolution of the chromosome which encompassed a ZNF gene cluster in a region of at least several tens of kb.

Genes from both KOX ZNF gene clusters are expressed

The cDNA clones KOX2, KOX31 and KOX21 have been completely sequenced: KOX2 maintains an open reading frame throughout its length (1136 bp) which includes four complete ZNF domains and half of a fifth, represented in the 3′ half of the cDNA (Figure 2a); KOX31 (972 bp) contains an open reading frame from its 5′ end which encodes five ZNF domains before terminating (Figure 2b); KOX21 (1007 bp) also contains an open reading frame from its 5′ end which encodes eight ZNF domains before terminating (accession number X69115). To determine the relationship between the cDNAs and their corresponding genomic loci, portions of the *ZNF11*, *ZNF33* and *ZNF37* sequences from both YACs yA4C5 and yA136E7 were amplified by polymerase chain reaction (PCR) and sequenced directly. For comparison, the cDNAs were also amplified and sequenced in the same experiments. In each case, PCR primers were chosen which gave products of identical size from both the cDNA from which they were derived and from genomic or YAC DNA. This avoided the problem of amplification of introns.

For the KOX2-related sequences, a PCR fragment of 977 bp was generated from all three DNA sources (corresponding to positions 110−1086 of the KOX2 cDNA; Figure 2a). Sequence comparison of PCR fragments was carried out over 669 bp (positions 193−861; Figure 2a) using four specific sequencing primers (see Materials and Methods). There was an absolute correspondence of KOX2 cDNA and yA136E7 sequences, demonstrating that the *ZNF11B* gene in 10q11.2 is expressed. The *ZNF11A* sequence, as amplified from yA4C5, contains base changes at 26 positions (3.9%), which result in 18 amino acid replacements (8.1%). This confirms that, although *ZNF11A* and *ZNF11B* are closely related, they are clearly different genes.

For the KOX31-related sequences, a PCR fragment of 850 bp was generated from all three templates (corresponding to positions 55−904 of the KOX31 cDNA; Figure 2b). Direct sequencing of PCR fragments over 255 bp (positions 348−602; Figure 2b) demonstrated identity of KOX31 cDNA and yA4C5 sequences, and therefore that *ZNF33A* is expressed. The *ZNF33B* sequence amplified from yA136E7 differed at 10 positions and also contained a single base insertion (position 439; Figure 2b). This changes the reading frame and truncates the corresponding protein, relative to that encoded by *ZNF33A*, due to a stop codon shortly after the insertion.

For the KOX21-related sequences, a PCR fragment of 919 bp was generated from the cDNA and yA4C5 (corresponding to positions 56−974 of the KOX21 cDNA; accession number X69115). Sequence comparison of PCR fragments was carried out over 496 bp (positions 144−639) using specific sequencing primers. KOX21 cDNA and yA4C5 sequences are identical, demonstrating that the *ZNF37A* gene in 10p11.2 is expressed. The *ZNF37B* sequence from yA136E7 was not sequenced, but presumably would represent a variant of that seen in yA4C5.

The KOX ZNF gene cluster duplication occurred before the divergence of human and great ape lineages

The above experiments demonstrated that both ZNF clusters contain similar but distinct genes, and that both loci contain expressed sequences. The existence of two clusters of related genes implies a duplication event covering an ancestral cluster: the degree of divergence of the comparable related sequences in each cluster can be used to estimate the evolutionary distance of the duplication. Using the alignments shown in Figure 2, the divergence between the *ZNF11A* and *ZNF11B*, and between the *ZNF33A* and *ZNF33B*, coding regions was calculated (according to the two-parameter model of ref. 21) to be ~4% in each case.

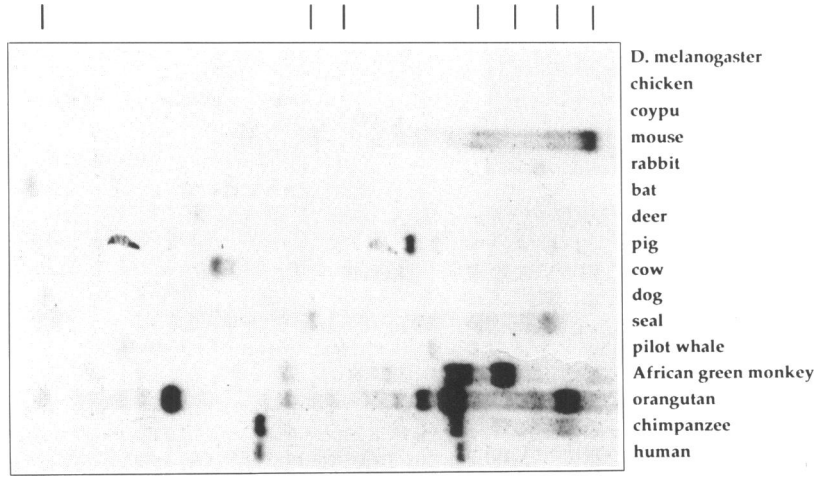

Figure 3. Zoo blot hybridised with KOX2 cDNA. EcoRI-digested DNAs were hybridised with KOX2 and washed to high stringency (0.2×SSC, 0.1% SDS, 65°C). Under these conditions, only KOX2-related sequences hybridise, with no cross-hybridisation to other ZNF genes observed (inferred from human pattern).

This figure is greater than the overall sequence divergence between man and the great apes (32,33), which suggests that the cluster duplication event occurred before these species diverged.

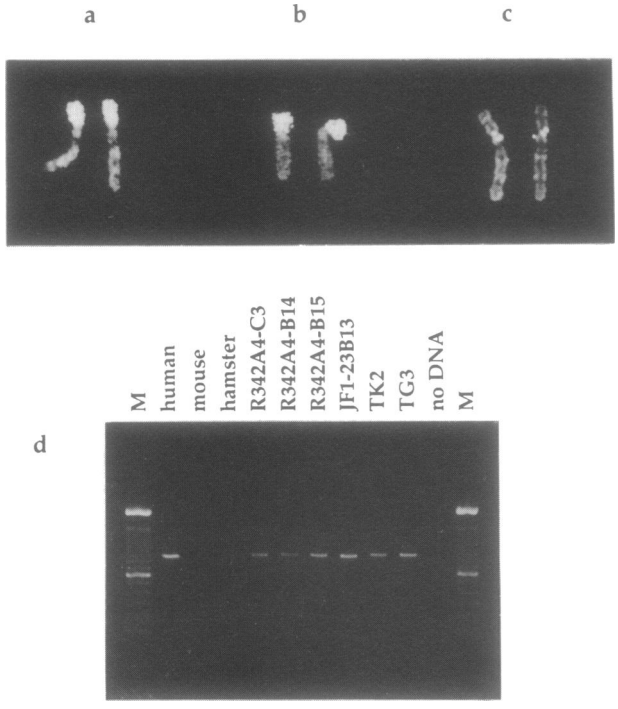

Figure 4. Localisation of *ZNF33A* to 10p11.2. **(a)** and **(b)** are chromosome paints where labelled Alu-PCR fragments from 10p-only hybrids R342A4-B14 and R342A4-B15 have been hybridised to normal human metaphase spreads. The chromosome 10 homologues have been shown in each case. **(c)** shows a FISH experiment demonstrating cross-hybridisation of yA136E7 (10q11.2) Alu-PCR fragments with the related ZNF locus on 10p11.2 of both chromosome 10 homologues. **(d)** is a PCR experiment where primers amplifying sequences from both *ZNF33* loci are used to demonstrate the presence of KOX31-related sequences in hybrids R342A4-B14, R342A4B15, R342A4C3 and JF1-23B13, and hence on 10p, since these hybrids contain no 10q sequences.

It might therefore be expected that sequences corresponding to *ZNF11* would also be duplicated in the apes, but not necessarily in species more distantly related to man. To test this, EcoRI digests of DNA from a number of species were probed with the KOX2 cDNA under stringent hybridisation conditions (Figure 3). As before, two EcoRI fragments are detected in human DNA, corresponding to *ZNF11A* (4.0 kb) and *ZNF11B* (1.75 kb). Remarkably, exactly the same pattern was seen with chimpanzee DNA, indicating not only a duplication of *ZNF11* genes in this ape but also a conservation of sequence at all four EcoRI sites defined in man. Multiple bands were also clearly seen in orangutan (*Pongo pygmaeus*) and the old world monkey *Ceropithecus aethiops*, but not in more distantly related species, such as pig or cow, where only a single hybridising band is seen. It can be concluded from these data that a *ZNF11* gene duplication event occurred relatively recently in the evolution of the human lineage. It seems certain that this took place at least before the divergence with chimpanzee, and since multiple KOX2-hybridising bands are also present in other apes and monkeys, the duplication event may have been prior to the branch points of most simian primates. However, duplication events, independent of that in the human-chimpanzee ancestor, cannot be ruled out in these other primate lineages.

The two related KOX ZNF loci are located on opposite sides of the chromosome 10 centromere

Although the KOX ZNF locus defined by YAC yA136E7 (cluster B) has been located definitively in band 10q11.2 (5), the exact band location of the related locus defined by yA4C5 (cluster A) was not known. *ZNF11A* was mapped previously to the interval 10p11.2−q11.2, above the translocation breakpoint in 10q11.2 defined in the TG3 and TK2 somatic cell hybrids (5). This breakpoint probably occurs within a large block of simple sequence satellite III DNA located close to the centromere (34). This suggests that *ZNF11A* and associated sequences either lie in close proximity to the centromere on 10q, amongst highly repetitive satellite DNA families, or that they are located on the short arm of chromosome 10. Either location would be surprising, since the simplest model of a ZNF gene cluster duplication would

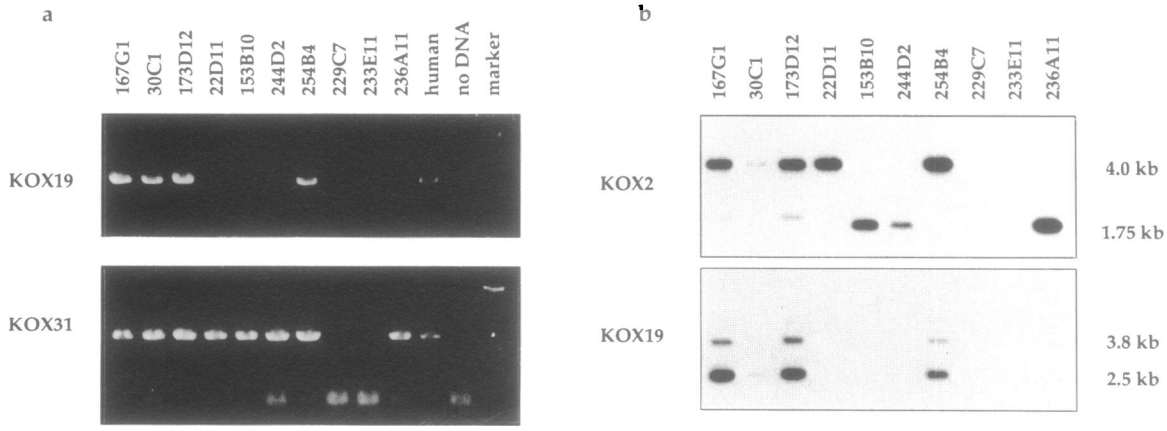

Figure 5. Physical linkage of *ZNF25* (KOX19) with cluster A genes in YACs. **(a)** PCR experiments detecting KOX19 (*ZNF25*) and KOX31 (*ZNF33A* or *ZNF33B*) sequences in CEPH YACs. Markers are a 100 bp ladder (GibcoBRL). **(b)** Filter hybridisations of CEPH YACs with KOX2 and KOX19 cDNA probes. Band sizes are indicated. Four YACs (167G1, 30C1, 173D12 and 254B4) contain *ZNF25* (KOX19). These YACs also contain *ZNF11A* and *ZNF33A* from 10p11.2, as does clone 22D11. Three YACs (153B10, 244D2 and 236A11) contain *ZNF11B* and *ZNF33B* from 10q11.2. YACs 229C7 and 223E11 do not contain DNA from this region of the genome.

place the genes of cluster A contiguous to the highly related sequences of cluster B in 10q11.2. The map location of cluster A was therefore determined more precisely using somatic cell hybrids.

A number of human-hamster somatic cell hybrid subclones were derived from a parent clone R342A4 (23) which contains chromosome 10 as its sole human genetic component. Three of these subclones (R342A4-B14, R342A4-B15, and R342A4-C3) were found to contain only the short arm of chromosome 10, as shown by chromosome painting (Figure 4a, b) of normal human metaphase spreads with DNA fragments derived from the hybrids by Alu-PCR (19). The 10p-containing hybrids were tested for the presence of *ZNF33* sequences by PCR using primers 4113 and 4114, which allow amplification from both *ZNF33* loci, as demonstrated by the results of Figure 2. A PCR fragment was obtained with all three hybrids (Figure 4d), demonstrating the presence of *ZNF33* sequences on the chromosome 10 short arm. A result consistent with this was obtained with a further unrelated somatic cell hybrid, JF1-23B13, which contains a t(10;21) translocation chromosome with 10pter-p11.2 as its only portion of chromosome 10 (25): JF1-23B13 is also positive for *ZNF33* sequences (Figure 4d). In both types of hybrid, these sequences must represent the *ZNF33A* locus, whose map position, along with those of *ZNF11A* and *ZNF37A* which form KOX ZNF cluster A, can now be localised to 10p11.2.

Fluorescence in situ hybridisation (FISH) experiments performed with Alu-PCR fragments from yA4C5 and yA136E7 confirmed the presence of distinct ZNF loci in 10p11.2 and 10q11.2. Indeed, in some FISH experiments, cross-hybridisation of the related sets of ZNF sequences flanking the centromere was seen. This is visible in Figure 4c, where a probe prepared from yA136E7 gave a signal at its cognate locus on 10q11.2, but also on 10p11.2.

Physical linkage of KOX ZNF gene cluster A to the ZNF25 (KOX19) gene

Provisional PFGE data had indicated that sequences from KOX ZNF gene cluster A might be physically linked to the *ZNF25* (KOX19) gene (L.Papi et al., unpublished data). Therefore, since YAC clone yA4C5, which contains the cluster A genes, did not contain *ZNF25* sequences, new YACs were isolated from the CEPH library by PCR screening with oligonucleotides derived from the KOX19 cDNA sequence. A library screen was also carried out by PCR with the KOX31-derived 4113 and 4114 primers, which should identify YACs containing either *ZNF33A* or *ZNF33B*. Eight YACs containing *ZNF33* sequences were isolated, and four of these were also positive in the *ZNF25* (KOX19) screen (Figure 5a). All eight YACs also contained KOX2-related sequences as determined by hybridisation of EcoRI-digested DNA with the KOX2 cDNA probe (Figure 5b). This is a further demonstration of close linkage of *ZNF11* and *ZNF33* sequences at each of the duplicated ZNF gene loci. Five of the eight YACs contain the 4.0 kb EcoRI KOX2-hybridising fragment indicative of *ZNF11A* in 10p11.2, and therefore contain cluster A sequences. The remaining three YACs contain the 1.75 kb EcoRI KOX2-hybridising fragment from the *ZNF11B* locus

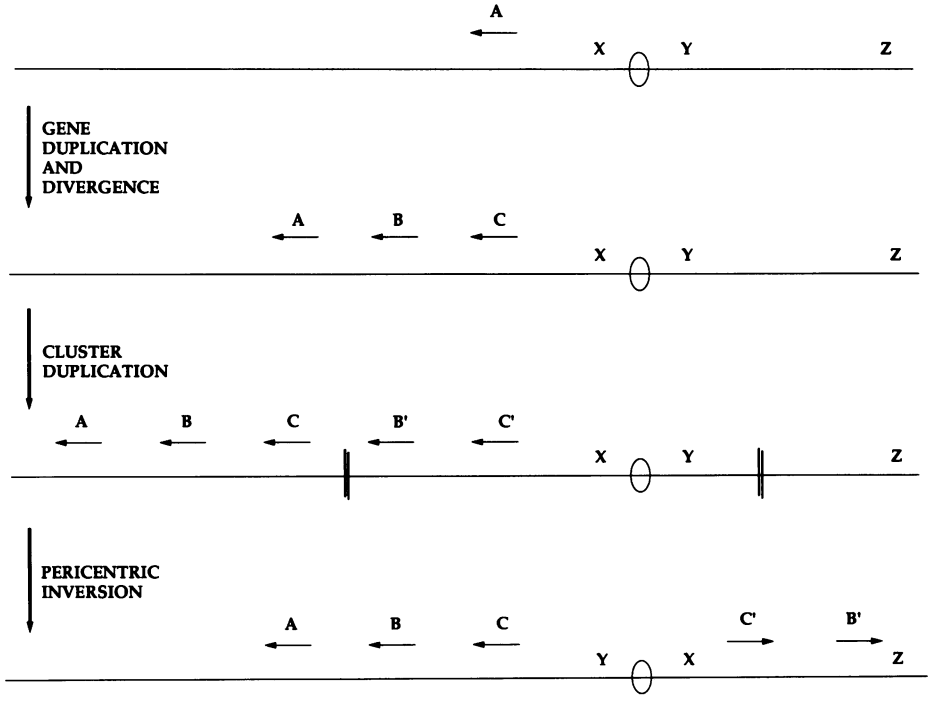

Figure 6. Model for evolution of the ZNF gene clusters in 10p11.2−q11.2. Three types of event are postulated: local gene duplications, which formed a cluster of related ZNF genes, A, B and C, after sequence divergence; a locus duplication, whereby at least part of the cluster of ZNF genes was duplicated on one chromosome arm (portrayed as genes B and C giving rise to their counterparts B′ and C′); and finally a pericentric inversion which separated the duplicated clusters, positioning them on opposite sides of the centromere. The breakpoints of the pericentric inversion are represented by twin vertical bars. X, Y and Z are sequences outside the duplicated region. As shown, the inversion breakpoint which disrupted the duplicated region coincides with a boundary of the cluster duplication event, but this need not have been the case. Only part of the duplicated region may have been transposed to the other chromosome arm.

in 10q11.2. Four of the five cluster A YACs are those which were also positive for *ZNF25* by PCR and by hybridisation with the KOX19 probe (Figure 5). That *ZNF25*, *ZNF11A* and *ZNF33A* sequences should all be present in four independently isolated YACs is strong evidence for linkage.

ZNF25 had previously been assigned to 10p11.2−q11.2 (5), but its physical linkage to cluster A genes refines its map location to 10p11.2. Consistent with this is the presence of *ZNF25* in somatic cell hybrids R342A4-B14 and JF1−23B13, which contain only the short arm of human chromosome 10 (data not shown). The KOX19 probe hybridises to two EcoRI fragments in both genomic DNA (5) and in YACs (Figure 5b). It is not yet clear whether this represents a *ZNF25* gene duplication, but since both EcoRI fragments are present within the same YAC clones, all KOX19-hybridising sequences are located in 10p11.2.

DISCUSSION

The ZNF genes corresponding to the KOX cDNA series (9) all encode ZNF proteins of the C_2H_2 Krüppel type (Class I, according to the nomenclature of ref. 35). In Drosophila, Class I ZNF genes are involved in body-plan formation (36−38) and the discovery of similar genes in higher organisms led to hypotheses that ZNF genes may also be important for mammalian development (39). Indeed, at least two examples of inherited syndromes with developmental defects are known to involve mutations in ZNF genes: GLI3 in Greig cephalopolydactyly syndrome (11) and WT1 in Denys−Drash syndrome (12). Other ZNF proteins have been shown to bind DNA at specific promoter sites (40,41), and they probably act by regulating gene activity (42−44). The KOX series of ZNF proteins therefore represent a set of potential gene regulators which deserve detailed characterisation.

One factor complicating the analysis of ZNF genes in mammals is their remarkably high number; several hundred are known to exist in man (8), and 60−70 of these are expressed in a single cell type (T cell; 9). Important for the identification of GLI3 and WT1 as candidate genes for their respective diseases was knowledge of their map location. We and others have therefore engaged on mapping projects which may give information on ZNF gene function and evolution. The results of these studies reveal a tendency for ZNF genes to map near fragile sites (45) or in subtelomeric locations (13). In addition, at the level of somatic cell hybrid and FISH mapping, they seem to cluster in certain regions of the genome, and one such cluster was proposed for human chromosome 10 (12) in the pericentromeric region (5). The molecular cloning data presented in this paper not only confirm the existence of clusters of related ZNF genes on chromosome 10, but also demonstrate duplication of an entire cluster: part of cluster A *(ZNF11A, ZNF33A, ZNF37A)* in 10p11.2, and cluster B *(ZNF11B, ZNF33B, ZNF37B)* in 10q11.2, obviously derive from an ancestral cluster. Of the other KOX-related ZNF genes in the chromosome 10 pericentromeric region, *ZNF22* in 10q11.2 (5) seems not to be closely linked to cluster B (this paper; and unpublished data). However, *ZNF25*, defined by cDNA clone KOX19, forms part of cluster A in 10p11.2, as demonstrated by YAC cloning and PFGE experiments. At present, it is not clear whether *ZNF25* is also duplicated. If it exists as a single copy in the genome, then the YACs linking *ZNF25* to the cluster A genes must contain one boundary of the cluster duplication event. [A gene *TCF8* encoding a ZNF protein transcription factor which inhibits expression of the interleukin-2 gene has recently been mapped to 10p11.2 (46), but its relationship to KOX-defined ZNF genes is at present unclear.] Thus most of the ZNF genes in this region of the genome may be organised in clusters.

Although not previously described in the literature at the molecular level, ZNF gene clustering is probably a common phenomenon in mammalian genomes. For example, mouse chromosome 17 contains a cluster of at least five ZNF genes in a region of approximately 500 kb (47; P.Little, personal communication). Clusters of related genes are thought to arise during evolution through duplication events, but clustering can also be exploited for coordinated regulation of the gene group, with the β-globin genes being the mammalian paradigm (48). It will therefore be of interest to determine whether the clustering of ZNF genes in 10p11.2−q11.2 is significant for their expression.

The location of homologous ZNF clusters on opposite sides of the centromere is unexpected since duplicated loci are usually contiguous in the genome. The duplication encompasses at least several tens of kilobases and is likely to extend beyond the immediate boundaries defined by the YAC clones yA4C5 and yA136E7. Other work in this laboratory has led to the isolation of further YACs containing duplicated sequences from 10p11.2−q11.2, including novel expressed genes (J.Kwok et al., unpublished data). Therefore, the duplication may extend over several hundred kilobases. As outlined in Figure 6, the most parsimonious model for the evolution of this region is a process involving first the build-up of a ZNF gene cluster by local gene duplication events, followed by a duplication of at least part of this cluster, and finally a pericentric inversion which distributed a portion of the duplicated region to the other chromosome arm. This sequence of events may represent a model for the dispersal of gene family members throughout the genome.

Sequence divergence and cross-species hybridisation studies suggest that the cluster duplication event occurred, and was fixed in the human lineage, before the separation from chimpanzee, and probably other great apes. Certainly this event postdates the mammalian radiation, since species more distantly related to man than the old world monkeys do not possess multiple copies of KOX2-related sequences. About the time of the pericentric inversion we can be less certain, although it must have occurred after the ZNF cluster duplication event. A pericentric inversion between 10p11.2 and 10q11.2 has not been described during human chromosome 10 evolution, but this would be difficult to detect cytogenetically. A similar inversion may have disrupted the immunoglobulin κ (IGK) gene complex on chromosome 2. Most IGK gene segments are located at 2p11−12, but some V_κ segments are found at 2q12−14, perhaps as a result of a pericentric inversion (49,50). Intriguingly, such an inversion has occurred during the recent evolution of human chromosome 2, being present in the chimpanzee precursor chromosome, but not in the related chromosomes of gorilla and orangutan (51). For the case of chromosome 10, in situ hybridisation studies on primate chromosomes with YACs from the duplicated region should provide information on the time of occurrence of the pericentric inversion.

Finally, the existence of a large duplicated region flanking the centromere of chromosome 10 has implications for genes located in 10p11.2−q11.2. If a gene maps within the duplication, it may well have a highly related counterpart on the opposite side of

the centromere. It will be intriguing to investigate whether these related genes, such as the KOX ZNF genes described here, have identical functions, or whether they have diverged sufficiently to develop distinct functional capacities.

ACKNOWLEDGEMENTS

We would like to thank Drs W.Amos and J.Pemberton for supplying some of the DNAs used in the zoo blot of Figure 3; Drs C.Todd and D.Bentley and Ms S.Hassock for help with the St Louis YAC library screening; Dr D.LePaslier and colleagues for secondary screens of the CEPH YAC library; Dr E.Holmes for calculating genetic distances between pairs of *ZNF11* and *ZNF33* sequences; and Dr P.Little for communicating unpublished results. This work was supported by a programme grant from the Cancer Research Campaign (CRC). A.T. was in receipt of an equipment grant from the Royal Society. L.P. was supported by the E.E.C. and the British Council. B.A.J.P. is a Gibb Fellow of the CRC. The Basel Institute for Immunology was founded and is supported by F.Hoffmann-La Roche Ltd.

REFERENCES

1. Nakamura,Y., Mathew,C.G.P., Sobol,H., Easton,D.F., Telenius,H., Bragg,T., Chin,K., Clark,J., Jones,C., Lenoir,G.M., White,R. and Ponder,B.A.J. (1989) *Genomics*, **5**, 199–203.
2. Wu,J., Carson,N.L., Myers,S., Pakstis,A.J., Kidd,J.R., Castiglione,C.M., Anderson,L.A., Hoyle,L.S., Genel,M., Verdy,M., Jackson,C.E., Simpson,N.E. and Kidd,K.K. (1990) *Am. J. Hum. Genet.*, **46**, 624–630.
3. Mole,S.E., Tunnacliffe,A., Mulligan,L.M., Cervini,R., Finocchiaro,G., Thiesen,H.-J., Nakamura,Y. and Ponder,B.A.J. (1991) *Cytogenet. Cell Genet.*, **58**, 1951–1952.
4. Rousseau-Merck,M.-F., Berger,R. and Thiesen,H.-J. (1991) *Cytogenet. Cell Genet.*, **58**, 1953.
5. Rousseau-Merck,M.-F., Tunnacliffe,A., Berger,R., Ponder,B.A.J. and Thiesen,H.-J. (1992) *Genomics*, **13**, 845–848.
6. Brown,R.S., Sander,C. and Argos,P. (1985) *FEBS. Lett.* , **186**, 271–274.
7. Miller,J., McLachlan,A.D. and Klug,A. (1985) *EMBO J.*, **4**, 1604–1614.
8. Bellefroid,E.J., Lecocq,P.J., Benhida,Q., Poncelet,D.A., Belayew,A. and Marial, J.A. (1989) *DNA*, **8**, 377–387.
9. Thiesen,H.-J. (1990) *New Biol.*, **2**, 363–374.
10. Pelletier,J., Bruening,W., Kashtan,C.E., Mauer,S.M., Manivel,J.C., Striegel,J.E., Houghton,D.C., Junien,C., Habib,R., Fouser,L., Fine,R.N., Silverman,B.L., Haber,D.A. and Housman,D. (1991) *Cell*, **67**, 437–447.
11. Vortkamp,A., Gessler,M. and Grzeschik,K.-H. (1991) *Nature*, **352**, 539–540.
12. Huebner,K., Druckt,T., Croce,C.M. and Thiesen,H.-J. (1991) *Am. J. Hum. Genet.*, **48**, 726–740.
13. Lichter,P., Bray,P., Ried,T., Dawid,I.B. and Ward,D.C. (1992) *Genomics*, **13**, 999–1007.
14. Hoovers,J.M.N., Mannens,M., John,R., Bliek,J., van Heyningen,V., Porteous,D.J., Leschot,N.J., Westerveld,A. and Little,P.F.R. (1992) *Genomics*, **12**, 254–263.
15. Brownstein,B.H., Silverman,G.A., Little,R.D., Burke,D.T., Korsmeyer,S.J., Schlessinger,D. and Olson,M.V. (1989) *Science*, **244**, 1348–1351.
16. Burke,D.T., Carle,G.F. and Olson,M.V. (1987) *Science*, **236**, 806–812.
17. Schwartz,D.C. and Cantor,C.R. (1984) *Cell*, **37**, 67–75.
18. Radice,P. and Tunnacliffe,A. (1992) *Genes Chrom. Cancer*, **5**, 50–56.
19. Nelson,D.L., Ledbetter,S.A., Corbo,L., Victoria,M.F., Ramirez-Solis,R., Webster,T.D., Ledbetter,D.H. and Caskey,C.T. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 6686–6690.
20. Hultman,T., Stahl,S., Hornes,E. and Uhlen,M. (1989) *Nucleic Acids Res.*., **17**, 4937–4946.
21. Kimura,M. (1980) *J. Mol. Evol.*, **16**, 111–120.
22. Mathew,C.G.P., Wakeling,W., Jones,E., Easton,D., Fisher,R., Strong,C., Smith,B., Chin,K., Little,K., Nakamura,Y., Shows,T.B., Jones,C., Goodfellow,P.J., Povey,S. and Ponder,B.A.J. (1990) *Ann. Hum. Genet.*, **54**, 121–129.
23. Trask,B.J., Ger van den Engh,M.C., Massa,H.F., Gray,J.W. and Van Dilla,M. (1991) *Somat. Cell Mol. Genet.*, **17**, 117–136.
24. Kao,F.T. and Puck,T.T. (1974) *Methods Cell Biol.*, **8**, 23–39.
25. Decker,R.A., Moore,J., Ponder,B. and Weber,J.L. (1992)*Genomics*, **12**, 604–606.
26. Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor.
27. Feinberg,A.P. and Vogelstein,B. (1983) *Anal. Biochem.*, **132**, 6–13.
28. Cremer,T., Lichter,P., Borden,J. and Ward,D.C. (1988) *Hum. Genet.*. 80, 235–246.
29. Suijkerbuijk,R.F., Matthopoulos,D., Kearney,L., Monard,S., Dhut,S., Cotter,F.E., Herbergs,J., Guerts van Kessel,A. and Young,B.D. (1992) *Genomics*, **13**, 355–362.
30. Pinkel,D., Straume,T. and Gray,J.W. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 2934–2938.
31. Fan,Y.S., Davis,L.M. and Shows,T.B. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 6223–6227.
32. Sibley,G. and Ahlquist,J.E. (1987) *J. Mol. Evol.*, **26**, 99–121.
33. Caccone,A. and Powell,J. (1989) *Evolution*, **43**, 925–942.
34. Jackson,M.S., Mole,S.E. and Ponder,B.A.J. (1992) *Nucleic Acids Res.*, **20**, 4781–4787.
35. Harrison,S.C. (1991) *Nature*, **353**, 715–719.
36. Rosenberg,U.B., Schröder,C., Priess,A., Kienlin,A., Cote,S., Riede,I. and Jäckle,H. (1986) *Nature*, **319**, 336–339.
37. Pankratz,M.J. and Jäckle,H. (1990) *Trends Genet.*, **6**, 287–292.
38. Kasai,Y., Nambu,J.R., Leiberman,P.M. and Crews,S.T. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 3414–3418.
39. Chowdhury,K., Deutsch,U. and Gruss,P. (1987) *Cell*, **48**, 771–778.
40. Stanojevic,D., Hoey,T. and Levine,M. (1989) *Nature*, **341**, 331–335.
41. Hsu,T., Gogos,J.A., Kirsh,S.A. and Kafatos,F.C. (1992) *Science*, **257**, 1946–1950.
42. Shi,Y., Seto,E., Chang,L.-S. and Shenk,T. (1991) *Cell*, **67**, 377–388.
43. Hoch,M., Gerwin,N., Taubert,H. and Jäckle,H. (1992) *Science*, **256**, 94–97.
44. Drummond,I.A., Madden,S.L., Rohwer-Nutter,P., Bell,G.I., Sukhatme,V.P. and Rauscher,F. (1992) *Science*, **257**, 674–678.
45. Seite,P., Huebner,K., Rousseau-Merck,M.F., Berger,R. and Thiesen,H.J. (1991) *Hum. Genet.*, **86**, 585–590.
46. Williams,T.M., Montoya,G., Wu,Y., Eddy,R.L., Byers,M.G. and Shows,T.B. (1992) *Genomics*, **14**, 194–196.
47. Crossley,P.H. and Little,P.F.R. (1991)*Proc. Natl. Acad. Sci. USA*, **88**, 7923–7927.
48. Grosveld,F., Blom van Assendelft,G., Greaves,D.R. and Kollias,G. (1987) *Cell*, **51**, 975–985.
49. Huber,C., Thiebe,R., Hameister,H., Smola,H., Lotscher,E. and Zachau,H.G. (1990) *Nucleic Acids Res.*, **18**, 3475–3478.
50. Zimmer,F.-J., Hameister,H., Schek,H. and Zachau,H.G. (1990) *EMBO J.*, **9**, 1535–1542.
51. Yunis,J.J. and Prakesh,O. (1982) *Science*, **215**, 1525–1530.